

# Network-Wide Distributed Carrier Frequency Offsets Estimation and Compensation

Jian Du and Yik-Chung Wu

## Abstract

In this paper, we propose a fully distributed algorithm for frequency offsets estimation in decentralized systems. With the proposed algorithm, each node estimates its frequency offsets by local computations and limited exchange of information with its direct neighbors. Such algorithm does not require any centralized information processing or knowledge of global network topology. It is shown analytically that the proposed algorithm always converges to the optimal estimates regardless of network topology. Simulation results demonstrate the fast convergence of the algorithm and show that estimation mean-squared-error at each node touches the centralized Cramér-Rao bound within a few iterations of message exchange. Therefore, the proposed method has low overhead and is scalable with network size.

## Index Terms

Carrier frequency offsets (CFOs) estimation, heterogeneous networks, factor graph, convergence analyses.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong (e-mail: dujianeee@gmail.com, ycwu@eee.hku.hk).

## I. INTRODUCTION

In wireless communication systems, local oscillators are used in transceivers to generate carrier signals required for up-conversion and down-conversion. Ideally, carrier frequencies produced by oscillators of each transceiver pair should be the same. However, in practice, frequencies synthesized from independent oscillators will be different from each other due to variation of oscillator circuits. The received signal impaired by carrier frequency offsets (CFOs) between transmitter and receiver leads to a continuous rotation of symbol constellation, resulting in degradation of system capacity and bit error rate (BER) [1]–[4]. Consequently, carrier frequency synchronization has always been a momentous issue in communication systems.

As modern wireless environments become more heterogeneous and decentralized, mobile terminals in a network engage more and more cooperative communications and distributed computations [5], new scenarios that require multiple wireless units to synchronize with each other arise. For example,

**Distributed beamforming:** As shown in Fig. 1(a), to improve the range of communications and save battery power during the transmission, multiple mobile terminals form a virtual antenna array and cooperatively direct a beam in the desired direction of transmission [6], [7]. Since each source node in the distributed beamformer has an independent local oscillator, common carrier frequency among all transmitters is crucial to ensure that a beam is aimed in the desired direction.

**Multi-cell cooperation:** In fully frequency reuse cellular systems as depicted in Fig. 1(b), despite different users interfere with each other, multiple base stations could coordinate their coding and decoding. It was shown that such joint-processing significantly outperforms a network with individual cell processing [8], [9]. Yet, multiple base-stations cooperation requires frequency synchronization so that there is no CFO between each pair of communication link [10].

**Heterogenous Networks (HetNets):** HetNets have attracted much attention from both industry and academia in the past few years. As shown in Fig. 1(c), in a 3-tier HetNet, a mobile may wish to be associated with different tier base stations in the uplink and downlink to obtain optimal performance [11]–[13]. However, multi-tier cooperation is possible only when different tiers of networks are frequency-synchronized to each other.

The above examples of network-wide synchronization problem can be summarized and reduced to a multi-node communication systems as shown in Fig. 2. Despite relative CFO between each

pair of nodes can be optimally estimated by existing methods [14]–[23], network-wide CFOs correction is difficult since each node needs to synchronize with multiple neighboring nodes with different relative CFOs at the same time. Making the problem more challenging is the fact that synchronization should be accomplished by local operations without knowing the global network structure since users move around and join different parts of the network randomly.

Pioneering works for network CFOs correction have been proposed in [24], [25]. By gathering all the information in a central processing unit, CFOs are estimated at the receiver and then feedback to corresponding transmitters to adjust the offsets. These methods are centralized, which are not suitable for large-scale network. On the other hand, [26]–[28] investigated methods for frequency synchronization in distributed beamforming systems. However, these methods require the formation and maintenance of special network structures (e.g., tree structure in [26], ring structure in [27] and chain structure in [28]), thus suffer from large overhead and long delay, and are not scalable with network size. Recently, [29] proposed the distributed frequency-locked loops (D-FLL) to control and synchronize the carrier frequencies of autonomous nodes based on average consensus principle. Notwithstanding the distributed carrier frequency calibration advantage, fully distributed D-FLL approach suffers from very slow convergence rate as shown in [29]. Furthermore, the D-FLL algorithm is designed exclusively for single path channel. Even in a simple point-to-point case with multi-path channel, the D-FLL cannot be applied directly.

In this paper, we propose a network-wide fully distributed CFOs estimation and compensation method which only involves local processing and information exchange between neighboring nodes. There is no need to have a central hub that aggregates information and no knowledge about the global network topology is required. The frequency offset of each oscillator is estimated and corrected locally in each node. After synchronization, there is no frequency offset between any pair of nodes in the network. The proposed algorithm is scalable with network size, and robust to topology changes. The convergence of the proposed method is also formally proved.

The rest of this paper is organized as follows. System model is presented in Section II. Fully distributed frequency offsets estimation and correction based on belief propagation (BP) is derived in Section III. The convergence property of the proposed method is analyzed in Section IV. Simulation results are given in Section IV and, finally, conclusions are provided in Section V.

*Notations:* Boldface uppercase and lowercase letters will be used for matrices and vectors,

respectively.  $\mathbb{E}$  denotes the expectation over the random variables. Superscripts  $H$  and  $T$  denote Hermitian and transpose, respectively. The symbol  $\mathbf{I}_N$  represents the  $N \times N$  identity matrix, while  $\mathbf{1}_K$  is an all one  $K$  dimensional vector. The symbol  $\otimes$  denotes the Kronecker product and  $\odot$  denotes the Hadamard product. Notation  $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{R})$  stands for the probability density function (pdf) of a Gaussian random vector  $\mathbf{x}$  with mean  $\boldsymbol{\mu}$  and covariance matrix  $\mathbf{R}$ . The symbol  $\propto$  represents the linear scalar relationship between two real valued functions.  $\text{diag}\{[a_1, \dots, a_N]\}$  corresponds to an  $N \times N$  diagonal matrix with diagonal components  $a_1$  through  $a_N$ , while  $\text{blkdiag}\{[\mathbf{A}_1, \dots, \mathbf{A}_N]\}$  corresponds to a block diagonal matrix with  $\mathbf{A}_1$  through  $\mathbf{A}_N$  as diagonal blocks. For two matrices  $\mathbf{X}$  and  $\mathbf{Y}$ ,  $\mathbf{X} \succeq \mathbf{Y}$  means that  $\mathbf{X} - \mathbf{Y}$  is a positive semi-definite matrix.

## II. SYSTEM MODEL

We consider a network consisting of  $K$  nodes distributed in a field as shown in Fig. 2. The topology of the network is described by a communication graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  of order  $K$ , where  $\mathcal{V} = \{1, \dots, K\}$  is the set of graph vertexes, and  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  is the set of graph edges. In the example shown in Fig. 2, the vertices are depicted by circles and the edges by lines connecting these circles. The neighborhood of node  $i$  is the set of nodes  $\mathcal{I}(i) \subset \mathcal{V}$  defined as  $\mathcal{I}(i) \triangleq \{j \in \mathcal{V} | \{i, j\} \in \mathcal{E}\}$ , i.e., those nodes that are connected via a direct communication link to node  $i$ . It is also assumed that any two distinct nodes can communicate with each other through finite hops, such graph is named strongly connected graph.

In general, relative CFOs exist between any pair of neighboring nodes, and can be estimated by traditional CFOs estimation methods. Let nodes  $i$  and  $j$  equipped with  $N_i$  and  $N_j$  antennas, respectively. Denote the frequency offsets (with respect to a reference frequency) of the  $q^{th}$  antenna on node  $i$  as  $\omega_q^i$ , while that of  $k^{th}$  antenna of node  $j$  as  $\omega_k^j$ . Then, the relative CFO between the  $q^{th}$  and  $k^{th}$  antenna of node  $i$  and  $j$  respectively is  $\epsilon_{q,k}^{i,j} \triangleq \omega_q^i - \omega_k^j$ . Here we consider the general case where each antenna can be associated with separate oscillator circuit. Therefore, for the Multiple Input Multiple Output (MIMO) system between node  $i$  and node  $j$ , there are  $N_i N_j$  relative CFOs denoted as  $\boldsymbol{\epsilon}^{i,j} \triangleq [\epsilon_{1,1}^{i,j}, \dots, \epsilon_{N_i,1}^{i,j}, \dots, \epsilon_{1,N_j}^{i,j}, \dots, \epsilon_{N_i,N_j}^{i,j}]^T$ . Such relative CFOs estimation in MIMO systems can be decomposed into  $N_j$  parallel Multiple Input Single Output (MISO) CFOs estimation problem [17]. For example, considering a flat-fading MISO system,

for the  $k^{th}$  receive antenna of node  $j$ , the received signal can be written as

$$y_k^{i,j}(t) = \sum_{q=1}^{N_i} h_{q,k}^{i,j} e^{j\epsilon_{q,k}^{i,j} t} z_q^i(t) + \xi_k^j(t) \quad t = 1, \dots, N, \quad (1)$$

where  $h_{q,k}^{i,j}$  are the unknown channel gains between the  $q^{th}$  antenna of node  $i$  and  $k^{th}$  antenna of node  $j$ ;  $j \triangleq \sqrt{-1}$ ;  $\{z_q^i(t)\}_{t=1}^N$  is the training sequence transmitted from the  $q^{th}$  antennas of node  $i$ ; and  $\xi_k^j(t)$  is the observation noise at the  $k^{th}$  antenna of node  $j$ . By stacking (1) with  $t = 1, \dots, N$  in vector form and omitting superscript  $i, j$  without confusion, the received vector  $\mathbf{y}_k \triangleq [y_k(1), \dots, y_k(N)]^T$  can be written as

$$\mathbf{y}_k = \mathbf{I}_k(\epsilon_k) \odot \mathbf{Z}_k \mathbf{h}_k + \boldsymbol{\xi}_k \quad k = 1, \dots, N_j, \quad (2)$$

where  $\mathbf{I}_k(\epsilon_k)$  is an  $N$ -by- $N_i$  Vandermonde matrix with its  $t^{th}$  row given by  $[e^{jt\epsilon_{1,k}}, e^{jt\epsilon_{2,k}}, \dots, e^{jt\epsilon_{N_i,k}}]$ ;  $\mathbf{Z}_k$  is the  $N$ -by- $N_i$  training sequence matrix with its  $t^{th}$  row  $[z_1(t), z_2(t), \dots, z_{N_i}(t)]$ ; and  $\boldsymbol{\xi}_k = [\xi_k(1), \dots, \xi_k(N)]^T$  is the observation noise. The parameters  $\epsilon_k \triangleq [\epsilon_{1,k}, \epsilon_{2,k}, \dots, \epsilon_{N_i,k}]^T$  and  $\mathbf{h}_k \triangleq [h_{1,k}, \dots, h_{N_i,k}]^T$  are the parameters need to be estimated.

If the noise is white and Gaussian, i.e.,  $\boldsymbol{\xi}_k \sim \mathcal{CN}(\boldsymbol{\xi}_k; \mathbf{0}, \sigma_k^2 \mathbf{I}_N)$ , joint relative CFOs and channels estimation have been extensively studied in the past two decades and the optimal estimates  $\hat{\epsilon}_k$  and  $\hat{\mathbf{h}}_k$  have been proposed in [16]–[20], with the mean-square-errors (MSEs) approaching the corresponding Cramér-Rao bounds (CRBs) in medium and high signal-to-noise ratio (SNR) ranges. From (2), the CRB of  $\epsilon_k$  can be shown to be [17]

$$\mathbf{B}_{\epsilon_k}(\epsilon_k, \mathbf{h}_k) = \frac{\sigma_k^2}{2} \{ \text{Re}[\mathbf{V}_k - \mathbf{T}_k^H (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{T}_k] \}^{-1}, \quad (3)$$

where  $\mathbf{V}_k \triangleq \text{diag}\{\mathbf{h}_k\} \mathbf{A}_k^H \mathbf{D}^2 \mathbf{A}_k \text{diag}\{\mathbf{h}_k\}$ ,  $\mathbf{T}_k \triangleq \mathbf{A}_k^H \mathbf{D} \mathbf{A}_k \text{diag}\{\mathbf{h}_k\}$ , with  $\mathbf{A}_k \triangleq \mathbf{I}_k(\epsilon_k) \odot \mathbf{Z}_k$  and  $\mathbf{D} \triangleq \text{diag}\{[1, 2, \dots, N]\}$ . Since there are  $N_j$  independent MISO estimation problems as in (2), the CRB for frequency estimation in MIMO system between node  $i$  and  $j$  is given by  $\mathbf{B}_{\epsilon}^{\{i,j\}}(\{\epsilon_k\}_{k=1}^{N_j}, \{\mathbf{h}_k\}_{k=1}^{N_j}) = \text{blkdiag}\{[\mathbf{B}_{\epsilon_1}(\epsilon_1, \mathbf{h}_1), \dots, \mathbf{B}_{\epsilon_{N_j}}(\epsilon_{N_j}, \mathbf{h}_{N_j})]\}$ .

After joint estimation of relative CFOs and channels, the relative CFOs between node  $i$  and  $j$  can be obtained as

$$\mathbf{r}_{i,j} = \mathbf{A}_{i,j} \boldsymbol{\omega}_i + \mathbf{A}_{j,i} \boldsymbol{\omega}_j + \mathbf{n}_{i,j}, \quad (4)$$

where  $\mathbf{r}_{i,j} \triangleq [\hat{\epsilon}_1^T, \hat{\epsilon}_2^T, \dots, \hat{\epsilon}_{N_j}^T]^T$  is the  $N_i N_j$  relative CFOs estimates;  $\mathbf{A}_{i,j} \triangleq \mathbf{I}_{N_i} \otimes \mathbf{1}_{N_j}$  and  $\mathbf{A}_{j,i} \triangleq -\mathbf{1}_{N_i} \otimes \mathbf{I}_{N_j}$ ; and  $\mathbf{n}_{i,j}$  is the estimation error. It is known that for the maximum likelihood (ML) estimates,  $\mathbf{r}_{i,j}$  is asymptotically Gaussian distributed with mean  $[\epsilon_1^T, \epsilon_2^T, \dots, \epsilon_{N_j}^T]^T =$

$\mathbf{A}_{i,j}\boldsymbol{\omega}_i + \mathbf{A}_{j,i}\boldsymbol{\omega}_j$  and covariance matrix equals to  $\mathbf{B}_\epsilon^{\{i,j\}}(\{\boldsymbol{\epsilon}_k\}_{k=1}^{N_j}, \{\mathbf{h}_k\}_{k=1}^{N_j})$  [30]. That is,  $\mathbf{r}_{i,j} \sim \mathcal{N}(\mathbf{r}_{i,j}; \boldsymbol{\epsilon}_{i,j}, \mathbf{B}_\epsilon^{\{i,j\}}(\{\boldsymbol{\epsilon}_k\}_{k=1}^{N_j}, \{\mathbf{h}_k\}_{k=1}^{N_j}))$ . Notice that the CRB depends on the true value of  $\{\boldsymbol{\epsilon}_k\}_{k=1}^{N_j}$  and  $\{\mathbf{h}_k\}_{k=1}^{N_j}$ , but since we have obtained the ML estimate  $\{\hat{\boldsymbol{\epsilon}}_k\}_{k=1}^{N_j}$  and  $\{\hat{\mathbf{h}}_k\}_{k=1}^{N_j}$ ,  $\mathbf{B}_\epsilon^{\{i,j\}}(\{\boldsymbol{\epsilon}_k\}_{k=1}^{N_j}, \{\mathbf{h}_k\}_{k=1}^{N_j})$  can be closely approximated by  $\mathbf{R}_{i,j} = \mathbf{B}_\epsilon^{\{i,j\}}(\{\hat{\boldsymbol{\epsilon}}_k\}_{k=1}^{N_j}, \{\hat{\mathbf{h}}_k\}_{k=1}^{N_j})$ .

With local information (4), the goal is to establish global frequency synchronization. That is, to estimate and compensate  $\boldsymbol{\omega}_i$  in each node based on estimation results of local relative CFOs  $\mathbf{r}_{i,j}$ .

*Remark 1:* The system model (2) can be extended to the cases where signals undergoing frequency selective fading channel and even doubly selective channel. Effective estimators have been extensively studied and MSE performance of these estimators were shown to touch the corresponding CRBs [31]–[33]. Thus, we can always establish the relative CFOs relationship as in (4).

*Remark 2:* After relative CFOs estimation, each receiver (node  $j$  in the example) obtains the estimate  $\mathbf{r}_{i,j}$  as well as the covariance matrix  $\mathbf{R}_{i,j}$ . By feeding back this information to the corresponding transmitter, node  $i$  also obtains the relative CFOs estimates and estimation error covariance.

### III. DISTRIBUTED CFOs ESTIMATION

#### A. Distributed CFOs Estimation via Belief Propagation

The optimal CFO estimator at each node in the Bayesian sense is the minimum mean square error (MMSE) estimator, which finds the mean of the marginalized posterior distribution:

$$\hat{\boldsymbol{\omega}}_i^{\text{MMSE}} \triangleq \int \cdots \int \boldsymbol{\omega}_i p(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_K | \{\mathbf{r}_{i,j}\}_{\{i,j\} \in \mathcal{E}}) d\boldsymbol{\omega}_2 \cdots d\boldsymbol{\omega}_K. \quad (5)$$

Here, without loss of generality, node 1 is assumed to be the reference node, so  $\boldsymbol{\omega}_1$  is not included in the marginalization. By using Bayes' theorem, the joint posterior distribution of all frequency offsets can be expressed as:

$$p(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_K | \{\mathbf{r}_{i,j}\}_{\{i,j\} \in \mathcal{E}}) \propto \prod_{i \in \mathcal{V}} p(\boldsymbol{\omega}_i) \prod_{\{i,j\} \in \mathcal{E}} p(\mathbf{r}_{i,j} | \boldsymbol{\omega}_i, \boldsymbol{\omega}_j), \quad (6)$$

where  $p(\boldsymbol{\omega}_i)$  is the prior distribution and  $p(\mathbf{r}_{i,j} | \boldsymbol{\omega}_i, \boldsymbol{\omega}_j) \sim \mathcal{N}(\mathbf{r}_{i,j}; \mathbf{A}_{i,j}\boldsymbol{\omega}_i + \mathbf{A}_{j,i}\boldsymbol{\omega}_j, \mathbf{R}_{i,j})$  is the likelihood function.

Notice that since the joint posterior distribution in (6) depends on interactions among all unknown variables, the computation of  $\hat{\omega}_i^{MMSE}$  in (5) requires gathering of all information in a central processing unit. However, such centralized processing is not favorable in large-scale networks.

In order to compute the optimal estimate (5) in a distributed way, one can exploit the conditional independence structure of the joint distribution (6), which is conveniently revealed by factor graph (FG). FG is an undirected bipartite graphical representation of a joint distribution that unifies direct and undirected graphical models. An example of FG in the context of network-wide synchronization is shown in Fig. 3. In the FG, there are two distinct kinds of nodes. One is variable nodes representing local synchronization parameters  $\omega_i$ . If there is a communication link between node  $i$  and node  $j$ , the corresponding variable nodes  $\omega_i$  and  $\omega_j$  are linked by the other kind of node, factor node  $f_{i,j} = p(\mathbf{r}_{i,j}|\omega_i, \omega_j)$  representing the local likelihood function <sup>1</sup>. On the other hand, the factor node  $f_i = p(\omega_i)$  denotes the prior distribution of frequency offsets of node  $i$ , and is connected only to the variable node  $\omega_i$ . Note that the FG is bipartite which means neighbors of a factor node must be variable nodes and vice versa.

From the FG, two kinds of messages are passed around: One is the message from factor node  $f$  (likelihood function  $f_{i,j}$  or prior distribution  $f_i$ ) to its neighboring variable node  $\omega_i$ , defined as the product of the function  $f$  with messages received from all neighboring variable nodes except  $\omega_i$ , and then marginalized for  $\omega_i$  [34]

$$m_{f \rightarrow i}^{(l)}(\omega_i) = \int \cdots \int f \times \prod_{\omega_j \in \mathcal{B}(f) \setminus \omega_i} m_{j \rightarrow f}^{(l-1)}(\omega_j) d\{\omega_j\}_{\omega_j \in \mathcal{B}(f) \setminus \omega_i}, \quad (7)$$

where  $\mathcal{B}(f)$  denotes the set of variable nodes that are direct neighbors of the factor nodes  $f$  on the FG and  $\mathcal{B}(f) \setminus \omega_i$  denotes the same set but with  $\omega_i$  removed. In (7),  $m_{j \rightarrow f}^{(l-1)}(\omega_j)$  is the other kind of message from variable node to factor node which is simply the product of the incoming messages on other links, i.e.,

$$m_{j \rightarrow f}^{(l)}(\omega_j) = \prod_{\tilde{f} \in \mathcal{B}(\omega_j) \setminus f} m_{\tilde{f} \rightarrow j}^{(l)}(\omega_j), \quad (8)$$

where  $\mathcal{B}(\omega_j)$  denotes the set of factor nodes that are direct neighbors of the variable nodes  $\omega_j$  on the FG. It can be seen from (7) and (8) that for both variable nodes and factor nodes,

<sup>1</sup>Note that  $f_{i,j}=f_{j,i}$ .

each outgoing message is a function of all incoming messages in the last round except the incoming message from the node where the outgoing message will be directed to. This essential restriction guarantees that for cycle-free FG, incoming and outgoing messages on each edge are independent, and at the end the algorithm produces correct marginal posterior distribution.

The two kinds of messages are iteratively updated at variable nodes and factor nodes, respectively. In any round of message exchange, a belief of  $\omega_i$  can be computed at variable node  $i$  as the product of all the incoming messages from neighboring factor nodes, which is given by

$$b^{(l)}(\omega_i) = \prod_{f \in \mathcal{B}(\omega_i)} m_{f \rightarrow i}^{(l)}(\omega_i). \quad (9)$$

Thereupon, the estimate of  $\omega_i$  in the  $l^{th}$  iteration is simply

$$\hat{\omega}_i^{(l)} = \int \omega_i b^{(l)}(\omega_i) d\omega_i. \quad (10)$$

Notice that after convergence, the belief  $b^{(l)}(\omega_i)$  at each variable node corresponds to the marginal distribution of that variable exactly when the underlying FG is loop free [34]. However, for the FG with loops, it is generally difficult to know if BP will converge [36]. Even if BP converge to a fixed point, there is no guarantee on the estimation accuracy. Despite the lack of general results on BP, in this paper, the convergence and optimality of BP for network-wide CFO estimation algorithm will be proved in section IV.

### B. Message Computation

In the BP framework, messages are passed and updated iteratively. In order to start the recursion, in the first round of message passing, it is reasonable to set the initial messages from factor nodes to variable nodes  $m_{f_i \rightarrow i}^{(0)}(\omega_i)$  and  $m_{f_{i,j} \rightarrow i}^{(0)}(\omega_i)$  as  $p(\omega_i)$  and non-informative message  $\mathcal{N}(\omega_i; \mathbf{v}_{f_{i,j} \rightarrow i}^{(0)}, \mathbf{C}_{f_{i,j} \rightarrow i}^{(0)})$ , respectively, where  $\mathbf{v}_{f_{i,j} \rightarrow i}^{(0)}$  can be arbitrarily chosen and  $[\mathbf{C}_{f_{i,j} \rightarrow i}^{(0)}]^{-1} = \mathbf{0}$ . Assuming  $p(\omega_i) = m_{f_{i,j} \rightarrow i}^{(0)}(\omega_i)$  is in Gaussian form  $\mathcal{N}(\omega_i; \mathbf{v}_i, \mathbf{C}_i)$  (if there is no prior information, we can set the mean to be zero and set the variance to be a large value, i.e., non-informative prior). Thereupon, based on the fact that the likelihood function  $f_{i,j}$  is also Gaussian, according to (7),  $m_{f_{i,j} \rightarrow i}^{(1)}(\omega_i)$  is a Gaussian function. In addition,  $m_{j \rightarrow f_{i,j}}^{(1)}(\omega_j)$  being the product of Gaussian functions in (8) is also a Gaussian function [38]. Thus during each round of message exchange, all the messages are Gaussian functions and only the mean vectors and covariance matrices need to be exchanged between factor nodes and variable nodes.



At this point, we can compute the messages at any iteration. In general, for the  $l^{th}$  ( $l = 2, 3, \dots$ ) round of message exchange, factor node  $f_{i,j}$  receive messages  $m_{j \rightarrow f_{i,j}}^{(l-1)}(\omega_j)$  from its neighboring variable nodes and then compute messages using (7). After some derivations, it can be obtained that

$$\begin{aligned} m_{f_{i,j} \rightarrow i}^{(l)}(\omega_i) &= \int p(\mathbf{A}_{i,j}, \mathbf{A}_{j,i} | \omega_i, \omega_j) m_{j \rightarrow f_{i,j}}^{(l-1)}(\omega_j) d\omega_j \\ &\propto \mathcal{N}(\omega_i; \mathbf{v}_{f_{i,j} \rightarrow i}^{(l)}, \mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}), \end{aligned} \quad (11)$$

where the inverse of covariance matrix is

$$[\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}]^{-1} = \mathbf{A}_{i,j}^T \left[ \mathbf{R}_{i,j} + \mathbf{A}_{j,i} \mathbf{C}_{j \rightarrow f_{i,j}}^{(l-1)} \mathbf{A}_{j,i}^T \right]^{-1} \mathbf{A}_{i,j}, \quad (12)$$

and the mean vector is

$$\mathbf{v}_{f_{i,j} \rightarrow i}^{(l)} = \mathbf{C}_{f_{i,j} \rightarrow i}^{(l)} \mathbf{A}_{i,j}^T \left[ \mathbf{R}_{i,j} + \mathbf{A}_{j,i} \mathbf{C}_{j \rightarrow f_{i,j}}^{(l-1)} \mathbf{A}_{j,i}^T \right]^{-1} (\mathbf{r}_{i,j} - \mathbf{A}_{j,i} \mathbf{v}_{j \rightarrow f_{i,j}}^{(l-1)}). \quad (13)$$

On the other hand, using (8), the messages passed from variable nodes to factor nodes can be computed as

$$\begin{aligned} m_{i \rightarrow f_{i,j}}^{(l)}(\omega_i) &= \prod_{f \in \mathcal{B}(\omega_i) \setminus f_{i,j}} m_{f \rightarrow i}^{(l)}(\omega_i) \\ &\propto \mathcal{N}(\omega_i; \mathbf{v}_{i \rightarrow f_{i,j}}^{(l)}, \mathbf{C}_{i \rightarrow f_{i,j}}^{(l)}), \end{aligned} \quad (14)$$

where

$$[\mathbf{C}_{i \rightarrow f_{i,j}}^{(l)}]^{-1} = \sum_{f \in \mathcal{B}(\omega_i) \setminus f_{i,j}} [\mathbf{C}_{f \rightarrow i}^{(l)}]^{-1}, \quad (15)$$

and

$$\mathbf{v}_{i \rightarrow f_{i,j}}^{(l)} = \mathbf{C}_{i \rightarrow f_{i,j}}^{(l)} \sum_{f \in \mathcal{B}(\omega_i) \setminus f_{i,j}} [\mathbf{C}_{f \rightarrow i}^{(l)}]^{-1} \mathbf{v}_{f \rightarrow i}^{(l)}. \quad (16)$$

Furthermore, during each round of message passing, each node can compute the belief for  $\omega_i$  using (9), which can be easily shown to be  $b_i^{(l)}(\omega_i) \sim \mathcal{N}(\omega_i; \boldsymbol{\mu}_i^{(l)}, \mathbf{P}_i^{(l)})$ , with the inverse of covariance matrix

$$[\mathbf{P}_i^{(l)}]^{-1} = \sum_{j \in \mathcal{I}(i)} [\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}]^{-1}, \quad (17)$$

and mean vector

$$\boldsymbol{\mu}_i^{(l)} = \mathbf{P}_i^{(l)} \sum_{j \in \mathcal{I}(i)} [\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}]^{-1} \mathbf{v}_{f_{i,j} \rightarrow i}^{(l)}. \quad (18)$$

When the algorithm converges or the maximum number of message exchange is reached, each node computes the CFOs according to (10) as

$$\hat{\omega}_i^{(l)} = \int \omega_i b^{(l)}(\omega_i) d\omega_i = \mu_i^{(l)}. \quad (19)$$

The iterative algorithm based on BP is summarized as follows. The algorithm is started by setting the message from factor node to variable node as  $m_{f_i \rightarrow i}^{(0)}(\omega_i) = p(\omega_i)$  and  $m_{f_{i,j} \rightarrow i}^{(0)}(\omega_i) = \mathcal{N}(\omega_i; \mathbf{v}_{f_{i,j} \rightarrow i}^{(0)}, \mathbf{C}_{f_{i,j} \rightarrow i}^{(0)})$  with  $\mathbf{v}_{f_{i,j} \rightarrow i}^{(0)} = \mathbf{0}$  and  $[\mathbf{C}_{f_{i,j} \rightarrow i}^{(0)}]^{-1} = \mathbf{0}$ . At each round of message exchange, every variable node computes the output messages to factor nodes according to (15) and (16). After receiving the messages from its neighboring variable nodes, each factor node computes its output messages according to (12) and (13). Such iteration is terminated when (18) converges (e.g., when  $\|\mu_i^{(l)} - \mu_i^{(l-1)}\| < \eta$ , where  $\eta$  is a threshold) or the maximum number of iteration is reached. Then the estimate of CFOs of each node is obtained as in (19).

*Remark 3:* In practical networks, there is neither factor nodes nor variable nodes. The two kinds of messages  $m_{i \rightarrow f_{i,j}}^{(l)}(\omega_i)$  and  $m_{f_{i,j} \rightarrow j}^{(l)}(\omega_j)$  are computed locally at node  $i$ , and only mean vector  $\mathbf{v}_{f_{i,j} \rightarrow j}^{(l)}(\omega_j)$  and covariance matrix  $\mathbf{C}_{f_{i,j} \rightarrow j}^{(l)}(\omega_j)$  are passed from node  $i$  to node  $j$  during each round of message exchange of BP. It can be seen the algorithm is fully distributed and each node only needs to exchange limited information with neighboring nodes.

*Remark 4:* Since each pair of node has knowledge of relative CFOs and channel between them, the BP message exchange can be performed as in point-to-point communications.

#### IV. THEORETICAL ANALYSES OF BP METHOD

It is generally known that if the FG contains cycles, such as the one shown in Fig. 3, messages can flow many times around the graph, leading to the possibility of divergence of BP algorithm [35]. A general sufficient condition for convergence of loopy FGs is given in [36]. Unfortunately, it requires the knowledge of the joint posterior distribution of all unknown variables as shown in (6), and is difficult to verify for large-scale dynamic networks. Recently, [37] proved the convergence of BP in the context of distributed clock offset synchronization in wireless sensor network. The convergence is established for scalar variables in which sub-stochastic and irreducible properties of BP message recursion were exploited. However, in vector variable case, the BP messages involve matrix inverses (see (12), (13), (15) and (16)), and the sub-stochastic and irreducible properties cannot be easily applied. In the following, we will prove

the convergence of BP messages in vector form, and show that the BP based CFO estimates asymptotically converge to the optimal MMSE solution regardless of network topology.

**Theorem 1.** The covariance matrix  $\mathbf{P}_i^{(l)}$  of belief  $b_i^{(l)}(\omega_i)$  at each node converges, and there exists a positive definite matrix  $\mathbf{P}_i^*$  such that  $\lim_{l \rightarrow +\infty} \mathbf{P}_i^{(l)} = \mathbf{P}_i^*$  regardless of network topology.

*Proof:* We begin with a few properties of positive semi-definite (p.s.d.) matrices and positive definite (p.d.) matrices. If  $\mathbf{X}$ ,  $\mathbf{Y}$ ,  $\mathbf{Z}$  are  $N_i$ -by- $N_i$  matrices and  $\mathbf{X} \succ \mathbf{0}$ ,  $\mathbf{Y} \succ \mathbf{0}$ ,  $\mathbf{Z} \succeq \mathbf{0}$ , then we have

Property i):  $\mathbf{X}^{-1} \succ \mathbf{0}$ .

Property ii):  $\mathbf{X} + \mathbf{Y} \succ \mathbf{0}$ .

Property iii):  $\mathbf{X} + \mathbf{Z} \succ \mathbf{0}$ .

Property iv):  $\mathbf{X} \succeq \mathbf{Y}$  if and only if  $\mathbf{Y}^{-1} \succeq \mathbf{X}^{-1}$  [39].

Property v):  $\mathbf{A}_{i,j}^T \mathbf{X} \mathbf{A}_{i,j} \succ \mathbf{0}$  and  $\mathbf{A}_{i,j}^T \mathbf{Z} \mathbf{A}_{i,j} \succeq \mathbf{0}$ , where  $\mathbf{A}_{i,j}$  is defined in (4).

Property vi):  $\mathbf{A}_{j,i} \mathbf{X} \mathbf{A}_{j,i}^T \succeq \mathbf{0}$ .

Properties i) to iv) are standard results in matrix analysis. Property v) is true due to the fact that  $\mathbf{A}_{i,j}$  is of full column rank. The proof of vi) follows from the definition of  $\mathbf{X} \succ \mathbf{0}$  which is  $\mathbf{y}^T \mathbf{X} \mathbf{y} \geq 0$  for any  $\mathbf{y}$  (including all zeros vector). The result is obtained if we let  $\mathbf{y} = \mathbf{A}_{j,i}^T \mathbf{x}$ .

Next, we investigate the updating properties of the message covariance matrix. Substituting (15) into (12), the covariance update rules from factor nodes to variable nodes are

$$[\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}]^{-1} = \mathbf{A}_{i,j}^T \left[ \mathbf{R}_{i,j} + \mathbf{A}_{j,i} \left[ \mathbf{C}_{f_j \rightarrow j}^{-1} + \sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [\mathbf{C}_{f_{k,j} \rightarrow j}^{(l-1)}]^{-1} \right]^{-1} \mathbf{A}_{j,i}^T \right]^{-1} \mathbf{A}_{i,j}. \quad (20)$$

From (20), we can deduce two consequences. First, if all message covariance  $\mathbf{C}_{f_{k,j} \rightarrow j}^{(l-1)}$  and prior covariance  $\mathbf{C}_{f_j \rightarrow j}$  on the right-hand-side of (20) are non-informative,  $[\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}]^{-1}$  cannot be updated, i.e.,  $[\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}]^{-1} = \mathbf{0}$ . On the other hand, if some of the  $[\mathbf{C}_{f_{k,j} \rightarrow j}^{(l-1)}]^{-1}$  or  $\mathbf{C}_{f_j \rightarrow j}^{-1}$  on the right-hand-side of (20) are p.d. while the remaining are  $\mathbf{0}$ , then  $[\mathbf{C}_{f_j \rightarrow j}^{-1} + \sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [\mathbf{C}_{f_{k,j} \rightarrow j}^{(l-1)}]^{-1}]^{-1} \succ \mathbf{0}$  according to property iii). Applying property vi), we have  $\mathbf{A}_{j,i} \left[ \mathbf{C}_{f_j \rightarrow j}^{-1} + \sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [\mathbf{C}_{f_{k,j} \rightarrow j}^{(l-1)}]^{-1} \right]^{-1} \mathbf{A}_{j,i}^T \succeq \mathbf{0}$ . Furthermore, since  $\mathbf{R}_{i,j}$  is the relative CFO estimation covariance, we have  $\mathbf{R}_{i,j} \succ \mathbf{0}$ . Thus,  $\mathbf{R}_{i,j} + \mathbf{A}_{j,i} \left[ \mathbf{C}_{f_j \rightarrow j}^{-1} + \sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [\mathbf{C}_{f_{k,j} \rightarrow j}^{(l-1)}]^{-1} \right]^{-1} \mathbf{A}_{j,i}^T \succ \mathbf{0}$ . Then, based on properties i) and v), we obtain  $[\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}]^{-1} \succ \mathbf{0}$ . We summarize the above discussion as

$$[\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}]^{-1} \begin{cases} = \mathbf{0}, & \text{if all } [\mathbf{C}_{f_{k,j} \rightarrow j}^{(l-1)}]^{-1} \text{ and } \mathbf{C}_{f_j \rightarrow j}^{-1} = \mathbf{0}, \\ \succ \mathbf{0}, & \text{if some of } [\mathbf{C}_{f_{k,j} \rightarrow j}^{(l-1)}]^{-1} \text{ or } \mathbf{C}_{f_j \rightarrow j}^{-1} \succ \mathbf{0}, \text{ while others are } \mathbf{0}. \end{cases} \quad (21)$$

Now we prove that for any node  $i$ , if there exists a directed path from node  $1 \rightarrow \dots \rightarrow j \rightarrow i$  in the network topology, there must be a finite iteration number  $s$  such that  $[C_{f_{i,j} \rightarrow i}^{(s+1)}]^{-1} \succeq [C_{f_{i,j} \rightarrow i}^{(s)}]^{-1} \succ [C_{f_{i,j} \rightarrow i}^{(s-1)}]^{-1} = \mathbf{0}$ . Initially, all  $[C_{f_{k,j} \rightarrow j}^{(0)}]^{-1}$  and  $C_{f_j \rightarrow j}^{-1}$  over the FG equal  $\mathbf{0}$  except  $C_{f_1 \rightarrow 1}^{-1} = \infty \mathbf{I}$  at the reference node. Hence, the message update starts from the reference node. More explicitly,  $\forall i: \{i, 1\} \in \mathcal{E}$ , the message covariance is obtained by putting  $j = 1$  into (20), which is

$$[C_{f_{i,1} \rightarrow i}^{(1)}]^{-1} = \mathbf{A}_{i,1}^T \mathbf{R}_{i,1}^{-1} \mathbf{A}_{i,1} \succ \mathbf{0}, \quad (22)$$

where the p.d. property is due to property v). Furthermore, since  $C_{f_1 \rightarrow 1}^{-1} = \infty \mathbf{I}$ , it will dominate the sum  $C_{f_j \rightarrow j}^{-1} + \sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [C_{f_{k,j} \rightarrow j}^{(l-1)}]^{-1}$  in (20), and lead to  $[C_{f_{i,1} \rightarrow i}^{(l)}]^{-1} = \mathbf{A}_{i,1}^T \mathbf{R}_{i,1}^{-1} \mathbf{A}_{i,1}$ . Thus, we have

$$\dots = [C_{f_{i,1} \rightarrow i}^{(l)}]^{-1} = \dots = [C_{f_{i,1} \rightarrow i}^{(2)}]^{-1} = [C_{f_{i,1} \rightarrow i}^{(1)}]^{-1} \succ [C_{f_{i,1} \rightarrow i}^{(0)}]^{-1} = \mathbf{0}. \quad (23)$$

Then, we consider all nodes  $i$  with a directed path node  $1 \rightarrow j \rightarrow i$ . In the  $1^{st}$  iteration, node  $j$  will take the position of node  $i$  in (23) implying  $[C_{f_{j,1} \rightarrow j}^{(1)}]^{-1} \succ \mathbf{0}$ , while  $[C_{f_{i,j} \rightarrow i}^{(1)}]^{-1}$  has not been updated in the first iteration, i.e.,  $[C_{f_{i,j} \rightarrow i}^{(1)}]^{-1} = \mathbf{0}$ . In the second iteration, from (21), we have  $[C_{f_{i,j} \rightarrow i}^{(2)}]^{-1} \succ \mathbf{0}$ . Since  $[C_{f_{k,j} \rightarrow j}^{(1)}]^{-1}$  and  $C_{f_j \rightarrow j}^{-1}$  equal  $\mathbf{0}$ , taking inverse on  $C_{f_j \rightarrow j}^{-1} + \sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [C_{f_{k,j} \rightarrow j}^{(1)}]^{-1}$  gives  $\left[ C_{f_j \rightarrow j}^{-1} + \sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [C_{f_{k,j} \rightarrow j}^{(1)}]^{-1} \right]^{-1} \succeq \left[ C_{f_j \rightarrow j}^{-1} + \sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [C_{f_{k,j} \rightarrow j}^{(2)}]^{-1} \right]^{-1}$ . Further applying properties vi), iv) and v), we obtain

$$\begin{aligned} & \underbrace{\mathbf{A}_{i,j}^T \left[ \mathbf{R}_{i,j} + \mathbf{A}_{j,i} \left[ C_{f_j \rightarrow j}^{-1} + \sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [C_{f_{k,j} \rightarrow j}^{(2)}]^{-1} \right]^{-1} \mathbf{A}_{j,i}^T \right]^{-1} \mathbf{A}_{i,j}}_{= [C_{f_{i,j} \rightarrow i}^{(3)}]^{-1}} \\ & \succeq \underbrace{\mathbf{A}_{i,j}^T \left[ \mathbf{R}_{i,j} + \mathbf{A}_{j,i} \left[ C_{f_j \rightarrow j}^{-1} + \sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [C_{f_{k,j} \rightarrow j}^{(1)}]^{-1} \right]^{-1} \mathbf{A}_{j,i}^T \right]^{-1} \mathbf{A}_{i,j}}_{= [C_{f_{i,j} \rightarrow i}^{(2)}]^{-1}}. \end{aligned} \quad (24)$$

Thus  $[C_{f_{i,j} \rightarrow i}^{(3)}]^{-1} \succeq [C_{f_{i,j} \rightarrow i}^{(2)}]^{-1} \succ [C_{f_{i,j} \rightarrow i}^{(1)}]^{-1} = \mathbf{0}$ . In general, for any node  $i$ , if there exists a directed path from node  $1 \rightarrow \dots \rightarrow j \rightarrow i$  in the network topology, there must be a finite iteration number  $s$  such that

$$[C_{f_{i,j} \rightarrow i}^{(s+1)}]^{-1} \succeq [C_{f_{i,j} \rightarrow i}^{(s)}]^{-1} \succ [C_{f_{i,j} \rightarrow i}^{(s-1)}]^{-1} = \mathbf{0}. \quad (25)$$

Finally, we divide the discussion into three cases, covering all possible relationships between two neighboring node  $i$  and  $j$ :

- a) there exists a path from node  $1 \rightarrow \dots \rightarrow j \rightarrow i$  and  $j \neq 1$ ;
- b) there exists a path from node  $1 \rightarrow i$ ;
- c) there is no path from node  $1 \rightarrow \dots \rightarrow j \rightarrow i$ .

For the first case, suppose  $[C_{f_{i,j} \rightarrow i}^{(l)}]^{-1} \succeq [C_{f_{i,j} \rightarrow i}^{(l-1)}]^{-1}$  holds for  $l > s$ . Since  $j \neq 1$ , there must be a node  $k$ , such that  $[C_{f_{k,j} \rightarrow j}^{(l)}]^{-1} \succeq [C_{f_{k,j} \rightarrow j}^{(l-1)}]^{-1}$ . Then, it can be easily shown that  $\sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [C_{f_{k,j} \rightarrow j}^{(l)}]^{-1} \succeq \sum_{f_{k,j} \in \mathcal{B}(\omega_j) \setminus f_{i,j}} [C_{f_{k,j} \rightarrow j}^{(l-1)}]^{-1}$ . Following the same arguments above (24), it can be obtained that  $[C_{f_{i,j} \rightarrow i}^{(l+1)}]^{-1} \succeq [C_{f_{i,j} \rightarrow i}^{(l)}]^{-1}$ . Hence, by induction we have

$$[C_{f_{i,j} \rightarrow i}^{(l)}]^{-1} \succeq \dots \succeq [C_{f_{i,j} \rightarrow i}^{(s+1)}]^{-1} \succeq [C_{f_{i,j} \rightarrow i}^{(s)}]^{-1} \succ \mathbf{0}, \quad \text{for } l > s. \quad (26)$$

For the second case, if there exists a path node  $1 \rightarrow i$ , the corresponding result is in (23). For the third case, if the path node  $1 \rightarrow \dots \rightarrow j \rightarrow i$  does not exist,  $[C_{f_{i,j} \rightarrow i}^{(l)}]^{-1}$  never get update, and always equals to  $[C_{f_{i,j} \rightarrow i}^{(0)}]^{-1} = \mathbf{0}$ .

Since strongly connected network is considered, there is at least one  $\tilde{j} \in \mathcal{I}(i)$  such that the first case is true, therefore, we obtain

$$\dots \succeq \sum_{j \in \mathcal{I}(i)} [C_{f_{i,j} \rightarrow i}^{(l+1)}]^{-1} \succeq \sum_{j \in \mathcal{I}(i)} [C_{f_{i,j} \rightarrow i}^{(l)}]^{-1} \succeq \dots \sum_{j \in \mathcal{I}(i)} [C_{f_{i,j} \rightarrow i}^{(s)}]^{-1} \succ \mathbf{0}, \quad \text{for } l > s. \quad (27)$$

Applying matrix inverse to (27) and using the definition of  $P_i^{(l)}$  in (17), we have

$$P_i^{(s)} \succeq \dots \succeq P_i^{(l)} \succeq P_i^{(l+1)} \succeq \dots \succ \mathbf{0}, \quad \text{for } l > s, \quad (28)$$

where the p.d. property of  $P_i^{(l)}$  is due to property i). Consequently such non-increasing p.d. matrix sequence converges to certain  $P_i^* \succ \mathbf{0}$  [40]. ■

The importance of Theorem 1 is that if a reference node exists, the belief covariance matrices always converge. Next, we investigate the convergence of belief mean vectors.

**Theorem 2.** The mean  $\mu_i^{(l)}$  of the belief  $b^{(l)}(\omega_i)$  converges to a fixed a vector  $\mu_i^*$  regardless of the network topology, i.e.,  $\lim_{l \rightarrow +\infty} \mu_i^{(l)} = \mu_i^*$ .

*Proof:* From the proof of Theorem 1, there are three cases of relationships between node  $i$  and node  $j$  (above (26)). For the first and second cases, the evolution of  $[C_{f_{i,j} \rightarrow i}^{(l)}]^{-1}$  are described by (23) and (26), respectively. Taking matrix inverse of (23) and (26), we can readily see that  $C_{f_{i,j} \rightarrow i}^{(l)}$  is a monotonically decreasing matrix sequence and bounded below by  $\mathbf{0}$ . Thus,  $C_{f_{i,j} \rightarrow i}^{(l)}$

is convergent. For the third case,  $\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}$  is never updated, and thus can also be viewed as convergent. On the other hand, computation of  $\mathbf{C}_{j \rightarrow f_{i,j}}^{(l)}$  depends on  $\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}$  as shown in (15). So, if  $\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}$  is convergent, then  $\mathbf{C}_{j \rightarrow f_{i,j}}^{(l)}$  is also convergent. In this proof, it is assumed that  $\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}$  and  $\mathbf{C}_{j \rightarrow f_{i,j}}^{(l)}$  have already converged to  $\mathbf{C}_{f_{i,j} \rightarrow i}^{(*)}$  and  $\mathbf{C}_{j \rightarrow f_{i,j}}^{(*)}$ , respectively, as the convergence of message covariance matrices do not depend on the message mean vectors.

Substituting (16) into (13), we obtain the mean update rules from factor nodes to variable nodes as

$$\mathbf{v}_{f_{i,j} \rightarrow i}^{(l)} = \underbrace{\mathbf{C}_{f_{i,j} \rightarrow i}^{(*)} \mathbf{A}_{i,j}^T \left[ \mathbf{R}_{i,j} + \mathbf{A}_{j,i} \mathbf{C}_{j \rightarrow f_{i,j}}^{(*)} \mathbf{A}_{j,i}^T \right]}_{\triangleq \mathbf{M}_{i,j}}^{-1} \left\{ \underbrace{\mathbf{r}_{i,j} - \left[ \mathbf{A}_{j,i} \mathbf{C}_{j \rightarrow f_{i,j}}^{(*)} \mathbf{C}_{f_{j \rightarrow j}}^{-1} \mathbf{v}_{f_{j \rightarrow j}} + \underbrace{\mathbf{A}_{j,i} \mathbf{C}_{j \rightarrow f_{i,j}}^{(*)}}_{\triangleq \mathbf{F}_{i,j}} \sum_{f_{j,k} \in \mathcal{B}(\omega_j) \setminus f_{j,i}} [\mathbf{C}_{f_{j,k} \rightarrow j}^{(*)}]^{-1} \mathbf{v}_{f_{j,k} \rightarrow j}^{(l-1)} \right]}_{\triangleq \mathbf{a}_{i,j}} \right\}. \quad (29)$$

Without loss of generality, define  $\mathbf{v}^{(l)}$  as a vector containing all  $\mathbf{v}_{f_{i,j} \rightarrow i}^{(l)}$  with ascending index<sup>2</sup> first on  $i$  and then on  $j$ . We can write (29) as

$$\mathbf{v}_{f_{i,j} \rightarrow i}^{(l)} = \mathbf{M}_{i,j} \mathbf{a}_{i,j} - \mathbf{M}_{i,j} \mathbf{F}_{i,j} \mathbf{\Gamma}_{i,j} \mathbf{v}^{(l-1)}, \quad (30)$$

where  $\mathbf{\Gamma}_{i,j}$  is a block matrix containing  $[\mathbf{C}_{f_{j,k} \rightarrow j}^{(*)}]^{-1}$  as component blocks such that (30) is satisfied. Stacking (30) for all  $i$  and  $j$ , and writing  $\mathbf{v}^{(l)} \triangleq [(\mathbf{v}_x^{(l)})^T, (\mathbf{v}_y^{(l)})^T]^T$ , where  $\mathbf{v}_x^{(l)}$  containing  $\mathbf{v}_{f_{i,j} \rightarrow i}^{(l)}$  with  $j = 1$ , while  $\mathbf{v}_y^{(l)}$  containing the remaining part of  $\mathbf{v}^{(l)}$ , we obtain

$$\underbrace{\begin{bmatrix} \mathbf{v}_x^{(l)} \\ \mathbf{v}_y^{(l)} \end{bmatrix}}_{\triangleq \mathbf{v}^{(l)}} = \underbrace{\begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix}}_{\triangleq \mathbf{Q}} \underbrace{\begin{bmatrix} \mathbf{v}_x^{(l-1)} \\ \mathbf{v}_y^{(l-1)} \end{bmatrix}}_{\triangleq \mathbf{v}^{(l-1)}} + \underbrace{\begin{bmatrix} \boldsymbol{\xi}_x \\ \boldsymbol{\xi}_y \end{bmatrix}}_{\triangleq \boldsymbol{\xi}}. \quad (31)$$

On the other hand, putting  $j = 1$  into (29) and notice that  $\mathbf{C}_{f_1 \rightarrow 1} = \mathbf{0}$  and  $\mathbf{C}_{1 \rightarrow f_{i,1}}^{(*)} = \mathbf{0}$  if  $j = 1$ , we have

$$\mathbf{v}_{f_{i,1} \rightarrow i}^{(l)} = [\mathbf{A}_{i,1}^T \mathbf{R}_{i,1}^{-1} \mathbf{A}_{i,1}]^{-1} \mathbf{A}_{i,1}^T \mathbf{R}_{i,1}^{-1} (\mathbf{r}_{i,1} - \mathbf{A}_{1,i} \boldsymbol{\omega}_1). \quad (32)$$

<sup>2</sup>The order of  $\mathbf{v}_{f_{i,j} \rightarrow i}^{(l)}$  arranged in  $\mathbf{v}^{(l)}$  in fact can be arbitrary as long as it does not change after the order is fixed.

which shows that  $\mathbf{v}_{f_{i,1} \rightarrow i}^{(l)}$  never changes with iteration number  $l$ . Since  $\mathbf{v}_x^{(l)}$  containing  $\mathbf{v}_{f_{i,1} \rightarrow i}^{(l)}$  as components,  $\mathbf{v}_x^{(l)}$  is fixed and independent of  $l$ . Hence, we can write (31) equivalently as

$$\underbrace{\begin{bmatrix} \mathbf{v}_{\text{fix}} \\ \mathbf{v}_y^{(l)} \end{bmatrix}}_{\mathbf{v}^{(l)}} = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix}}_{\mathbf{Q}} \underbrace{\begin{bmatrix} \mathbf{v}_{\text{fix}} \\ \mathbf{v}_y^{(l-1)} \end{bmatrix}}_{\mathbf{v}^{(l-1)}} + \underbrace{\begin{bmatrix} \mathbf{0} \\ \boldsymbol{\xi}_y \end{bmatrix}}_{\boldsymbol{\xi}} \quad (33)$$

where  $\mathbf{v}_{\text{fix}} = \mathbf{v}_x^{(l)}$  represents the stacked messages for  $j = 1$ . Notice that  $[\mathbf{Q}_1 \mathbf{Q}_2]$  depends on  $-\mathbf{M}_{i,j} \mathbf{F}_{i,j} \mathbf{\Gamma}_{i,j}$ , while  $\boldsymbol{\xi}_y$  depends on  $\mathbf{M}_{i,j} \mathbf{a}_{i,j}$ . It is obvious that  $\mathbf{Q}_1$ ,  $\mathbf{Q}_2$  and  $\boldsymbol{\xi}_y$  are independent of iteration number  $l$ . Next, we will show a property of  $\mathbf{Q}_2$ .

Since  $\mathbb{E}_{\boldsymbol{\omega}_i, \boldsymbol{\omega}_j, \mathbf{n}_{i,j}} \{\mathbf{r}_{i,j}\} = \mathbb{E}\{\mathbf{A}_{i,j} \boldsymbol{\omega}_i + \mathbf{A}_{j,i} \boldsymbol{\omega}_j + \mathbf{n}_{i,j}\} = \mathbf{0}$ , and  $[\mathbf{C}_{f_j \rightarrow j}]^{-1} = \mathbf{0}$  with  $j \neq 1$  for non-informative prior, taking expectation on both sides of (33), we have

$$\underbrace{\begin{bmatrix} \bar{\mathbf{v}}_{\text{fix}} \\ \bar{\mathbf{v}}_y^{(l)} \end{bmatrix}}_{\triangleq \bar{\mathbf{v}}^{(l)}} = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix}}_{\mathbf{Q}} \underbrace{\begin{bmatrix} \bar{\mathbf{v}}_{\text{fix}} \\ \bar{\mathbf{v}}_y^{(l-1)} \end{bmatrix}}_{\triangleq \bar{\mathbf{v}}^{(l-1)}} \quad (34)$$

or equivalently

$$\bar{\mathbf{v}}^{(l)} = \mathbf{Q}^l \bar{\mathbf{v}}^{(0)}, \quad (35)$$

where  $\bar{x}$  denotes the expectation of  $x$ . Since there is always a positive value  $\eta$ , satisfying  $\eta > \sum_{i \neq j} |[\mathbf{Q}]_{i,j}|$  for all  $i$ , we have  $\eta \mathbf{I} + \mathbf{Q}$  is strictly diagonally dominant and then  $\eta \mathbf{I} + \mathbf{Q}$  is nonsingular [41]. Hence, arbitrary initial value  $\bar{\mathbf{v}}^{(0)}$  can be expressed in terms of the eigenvectors of  $\eta \mathbf{I} + \mathbf{Q}$  as  $\bar{\mathbf{v}}^{(0)} = \sum_{d=1}^D c_d \mathbf{q}_d$ , where  $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_D$  are the eigenvectors of  $\eta \mathbf{I} + \mathbf{Q}$ . Since the eigenvectors of  $\eta \mathbf{I} + \mathbf{Q}$  is the same as that of  $\mathbf{Q}$ , and the eigenvalues of  $\eta \mathbf{I} + \mathbf{Q}$  are  $\eta + \lambda_d$  ( $1 \leq d \leq D$ ), where  $\lambda_d$  is the eigenvalue of  $\mathbf{Q}$ , we have

$$\bar{\mathbf{v}}^{(l)} = \mathbf{Q}^l \bar{\mathbf{v}}^{(0)} = \sum_{d=1}^D c_d \lambda_d^l \mathbf{q}_d. \quad (36)$$

Without loss of generality, suppose  $\lambda_d$  are arranged in descending order as below

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_D|. \quad (37)$$

Let the eigenvalue with the largest magnitude has a multiplicity of  $d_0$ . Then  $\lambda_d/\lambda_1 < 1$  for  $d > d_0$  and  $(\lambda_d/\lambda_1)^l = 0$  if  $l$  is large enough. We then obtain

$$\bar{\mathbf{v}}^{(l)} = \lambda_1^l \sum_{d=1}^{d_0} c_d \mathbf{q}_d, \quad (38)$$

for large  $l$ . Taking expectation on (32), we have  $\bar{\mathbf{v}}_{f_{i,1} \rightarrow i}^{(l)} = -[\mathbf{A}_{i,1}^T \mathbf{R}_{i,1}^{-1} \mathbf{A}_{i,1}]^{-1} \mathbf{A}_{i,1}^T \mathbf{R}_{i,1}^{-1} \mathbf{A}_{1,i} \boldsymbol{\omega}_1$ . It is obvious that  $\bar{\mathbf{v}}_{f_{i,1} \rightarrow i}^{(l)}$  never change with the iteration number  $l$ . Hence, the first element of  $\bar{\mathbf{v}}^{(l)}$  is  $\bar{\mathbf{v}}^{(l)}(1) \triangleq \gamma_c$ , which is a constant, and according to (38), we have  $\lambda_1^l = \frac{\gamma_c}{\sum_{d=1}^{d_0} c_d \mathbf{q}_d(1)}$  for  $l$  large enough. Substituting it back to (38) yields

$$\bar{\mathbf{v}}^{(l)} = \frac{\gamma_c \sum_{d=1}^{d_0} c_d \mathbf{q}_d}{\sum_{d=1}^{d_0} c_d \mathbf{q}_d(1)}. \quad (39)$$

It is obvious that  $\bar{\mathbf{v}}^{(l)}$  does not change when  $l$  is large enough, and therefore,  $\bar{\mathbf{v}}^{(l)}$  in (35) converges. Since (34) and (35) are equivalent,  $\bar{\mathbf{v}}_y^{(l)}$  in (34) also converges. With iteration equation in (34) rewritten as

$$\bar{\mathbf{v}}_y^{(l)} = \mathbf{Q}_2 \bar{\mathbf{v}}_y^{(l-1)} + \mathbf{Q}_1 \bar{\mathbf{v}}_{\text{fix}}, \quad (40)$$

and since (40) converges, the spectrum radius  $\rho(\mathbf{Q}_2) < 1$  [42].

Now rewriting (33) as

$$\mathbf{v}_y^{(l)} = \mathbf{Q}_2 \mathbf{v}_y^{(l-1)} + \mathbf{Q}_1 \mathbf{v}_{\text{fix}} + \boldsymbol{\xi}_y. \quad (41)$$

With  $\mathbf{Q}_1 \mathbf{v}_{\text{fix}}$  being a constant vector, and  $\rho(\mathbf{Q}_2) < 1$ , we also have (41) converges. Thus, the sequence  $\mathbf{v}_{f_{i,j} \rightarrow i}^{(l)}$  in (29) is convergent for any initial vectors  $\mathbf{v}_{f_{i,j} \rightarrow i}^{(0)}$  [42]. Finally, with  $\boldsymbol{\mu}_i^{(l)}$  defined in (18), since  $\mathbf{P}_i^{(l)}$ ,  $\mathbf{C}_{f_{i,j} \rightarrow i}^{(l)}$  and  $\mathbf{v}_{f_{i,j} \rightarrow i}^{(l)}$  converge, we can draw the conclusion that the vector sequence  $\{\boldsymbol{\mu}_i^{(1)}, \boldsymbol{\mu}_i^{(2)}, \dots\}$  converges. ■

Although Theorem 2 states that the proposed iterative estimation for  $\{\boldsymbol{\omega}_i\}_{i \in \mathcal{V}}$  converges to a fixed point  $\{\boldsymbol{\mu}_i^*\}_{i \in \mathcal{V}}$ , We still need to answer the important question that how accurate the converged  $\{\boldsymbol{\mu}_i^*\}_{i \in \mathcal{V}}$  is?

**Theorem 3.** The BP message mean vector  $\boldsymbol{\mu}_i^*$  converges to the optimal estimates  $\hat{\boldsymbol{\omega}}_i^{\text{MMSE}}$ . Furthermore, for the non-informative prior of  $\boldsymbol{\omega} = [\boldsymbol{\omega}_2^T, \dots, \boldsymbol{\omega}_K^T]^T$ , the estimation MSE of  $\boldsymbol{\mu}^* = [(\boldsymbol{\mu}_2^*)^T, \dots, (\boldsymbol{\mu}_K^*)^T]^T$  asymptotically (in high SNR or large training length  $N$  or both) equals the centralized CRB of  $\boldsymbol{\omega}$ :

$$\text{CRB}(\boldsymbol{\omega}) = (\mathbf{A}^T \mathbf{R}^{-1} \mathbf{A})^{-1}, \quad (42)$$

where  $\mathbf{A}$  is obtained from stacking (4) into the form of  $\mathbf{r} = \mathbf{A}\boldsymbol{\omega} + \mathbf{n}$ , with  $\mathbf{r}$  being a vector containing  $\mathbf{r}_{i,j}$  with ascending index first on  $i$  and then on  $j$ ;  $\mathbf{R}$  is a block diagonal matrix with  $\mathbf{R}_{i,j}$  as block diagonal and with the same order as  $\mathbf{r}_{i,j}$  in  $\mathbf{r}$ .



*Proof:* Since the joint posterior distribution in (6) is multivariate Gaussian and it is known that if Gaussian BP converges, the mean of the beliefs computed by BP equals the mean of the marginal posterior distribution in (5), i.e.,  $\boldsymbol{\mu}_i^* = \hat{\boldsymbol{\omega}}_i^{\text{MMSE}}$  [35].

Notice that stacking (5) into  $\hat{\boldsymbol{\omega}}^{\text{MMSE}} = [(\hat{\boldsymbol{\omega}}_2^{\text{MMSE}})^T, \dots, (\hat{\boldsymbol{\omega}}_K^{\text{MMSE}})^T]^T$  gives

$$\hat{\boldsymbol{\omega}}^{\text{MMSE}} = \int \dots \int \boldsymbol{\omega} p(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_K | \{\mathbf{r}_{i,j}\}_{\{i,j\} \in \mathcal{E}}) d\boldsymbol{\omega}_2 \dots d\boldsymbol{\omega}_K. \quad (43)$$

It is obvious that  $\boldsymbol{\mu}^* = [(\boldsymbol{\mu}_2^*)^T, \dots, (\boldsymbol{\mu}_K^*)^T]^T$  equals the centralized joint MMSE estimator  $\hat{\boldsymbol{\omega}}^{\text{MMSE}}$ . Putting (6) into (43) and in case of non-informative prior,  $\hat{\boldsymbol{\omega}}^{\text{MMSE}}$  is the mean of the joint likelihood function. Since the mean and maximum of a Gaussian distribution are the same, therefore,  $\boldsymbol{\mu}^*$  equals the centralized joint maximum likelihood estimator and it asymptotically approaches the centralized CRB [30]. Finally, stacking (4) into the form

$$\mathbf{r} = \mathbf{A}\boldsymbol{\omega} + \mathbf{n}, \quad (44)$$

where  $\mathbf{r}$  is a vector containing  $\mathbf{r}_{i,j}$  with ascending indexes first on  $i$  and then on  $j$ ; and  $\mathbf{n}$  containing  $\mathbf{n}_{i,j}$  with the indexes  $i, j$  ordered in the same way as in  $\mathbf{r}$ . Since  $\mathbf{n} \sim \mathcal{N}(\mathbf{n}; \mathbf{0}, \mathbf{R})$ , where  $\mathbf{R}$  is a block diagonal matrix with  $\mathbf{R}_{i,j}$  as block diagonal and with the same order as  $\mathbf{r}_{i,j}$  in  $\mathbf{r}$ , and (44) is a standard linear model, the CRB for  $\boldsymbol{\omega}$  is given by  $\text{CRB}(\boldsymbol{\omega}) = (\mathbf{A}^T \mathbf{R}^{-1} \mathbf{A})^{-1}$  [30]. ■

## V. SIMULATION RESULTS

This section presents numerical results to assess the performance of the proposed algorithm. Estimation MSE are presented for CFO estimation over the whole random network, which consists of 14 nodes randomly located in a  $[0, 100] \times [0, 100]$  area. The communication range for each node is 38. In each trial, CFO of each antenna on each node (except node 1 where CFO is zero) is generated independently and is uniformly distributed in the range  $2\pi[-0.05, 0.05]$ . It is assumed that we do not have prior information on the distribution of CFOs, so we set  $p(\boldsymbol{\omega}_i) = \mathcal{N}(\boldsymbol{\omega}_i; \mathbf{0}, +\infty \mathbf{I})$ ,  $i = 2, \dots, 14$ . Besides, the channel between each pair of nodes is Rayleigh flat-fading. The relative CFOs and channels are first estimated based on the algorithm in [19], with training length  $N$ . Then the BP algorithm is executed for network-wide CFOs estimation and compensation. 5000 simulation runs were performed to obtain the average performance for each point in the figures.

First, consider the network shown in Fig. 2 and each node equipped with two antennas. We employ training with length  $N = 16$  for relative CFOs estimation and the SNR during training stage in each node is 30dB. Fig. 4 shows the sum MSE<sup>3</sup> of  $\omega_i$  for  $i = \{3, 6, 2\}$  as a function of BP iteration number  $l$ . It can be seen that the MSEs decrease quickly and touch the corresponding CRBs in only a few iterations.

Fig. 5 shows the average sum MSE of  $\{\omega_i\}_{i \in \mathcal{V}}$  versus SNRs for different training length  $N$ . The network is randomly generated within the  $[0, 100] \times [0, 100]$  area in each trial, and each node is equipped with 2 antennas. As shown in the figure, the MSEs of proposed distributed algorithm achieve the best performance as the MSEs touch the corresponding CRBs. This verifies Theorem 3. Furthermore, with increasing  $N$ ,  $\mathbf{R}_{i,j}$  touches CRB of the relative CFO  $\mathbf{B}_\epsilon^{\{i,j\}}(\{\epsilon_k\}_{k=1}^{N_j}, \{\mathbf{h}_k\}_{k=1}^{N_j})$  at lower SNR, and thus the estimation MSEs of  $\omega_i$  achieves the corresponding CRBs earlier.

Finally, we compare the performance of the proposed algorithm with that of the D-FLL approach in [29], where CFOs are adjusted to an average common value in each iteration based on consensus principle. The estimation error of consensus method at the  $l^{th}$  iteration is measured by the total mean-square deviation of the individual variables from their average, which is

$$\text{MSE}_{\text{consensus}}^{(l)} = \frac{1}{K} \sum_{i=1}^K \mathbb{E} \left\{ \left\| \boldsymbol{\mu}_i^{(l)} - \frac{1}{K} \sum_{i=1}^K \boldsymbol{\mu}_i^{(l)} \right\|^2 \right\}. \quad (45)$$

On the other hand, the proposed algorithm estimates the absolute CFO values, therefore, the network estimation MSE at the  $l^{th}$  iteration is

$$\text{MSE}_{\text{BP}}^{(l)} = \frac{1}{K-1} \sum_{i=2}^K \mathbb{E} \left\{ \left\| \boldsymbol{\mu}_i^{(l)} - \omega_i \right\|^2 \right\}. \quad (46)$$

We consider the 14 nodes randomly located within the  $[0, 100] \times [0, 100]$  area and for fair comparison with D-FLL, each node is equipped with a single antenna. For consensus method, pilots of length 16 are transmitted by each node in each iteration. However, for the proposed method, 16 pilots are used only in the relative CFOs estimation at the initial phase. The convergence performance of the two algorithms at different SNRs are shown in Fig. 6. It is apparent that convergence speed of the consensus algorithm decreases with SNRs, and in general takes several hundreds of iterations to converge. For example, at SNR = 5dB, around

<sup>3</sup>Sum MSE over the two antennas.

800 iterations is required. Together with the fact that a training of length 16 is being transmitted by each node in each iteration, this is a huge burden to the network, and also causing long delay in the synchronization process. On the other hand, the proposed method requires only a few iterations of message exchange of two real numbers (mean and variance) to approach the corresponding CRBs.

## VI. CONCLUSIONS

In this paper, a fully distributed CFOs estimation algorithm for cooperative and distributed networks was proposed. The algorithm is based on BP and is easy to be implemented by exchanging limited amount of information between neighboring nodes. Therefore, it has low overhead and is scalable with network size. Furthermore, it was shown analytically that the proposed distributed algorithm converges to the optimal solution with estimation MSE coincides with the centralized CRB asymptotically regardless of the network topology. Simulation results showed that the MSE of the proposed method touches the CRB within only a few iterations.

## REFERENCES

- [1] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing*. Wiley-InterScience, John Wiley & Sons, Inc., 1997.
- [2] T. Pollet, M. V. Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," *IEEE Trans. Commun.*, vol. 43, pp. 191–193, Feb./Mar./Apr. 1995.
- [3] H. Steendam and M. Moeneclaey, "Synchronization sensitivity of multicarrier systems," *European Commun.*, ETT special issue on multi-carrier spread spectrum, vol. 52, no. 5, pp. 834–844, May 2004.
- [4] V. K. V. Gottumukkala and H. Minn, "Capacity analysis and pilotdata power allocation for MIMO-OFDM with transmitter and receiver IQ imbalances and residual carrier frequency offset," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 553–565, Feb. 2012.
- [5] I.-H. Wang and D. N. C. Tse, "Interference mitigation through limited receiver cooperation," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2913–2940, May 2011.
- [6] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z. Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4306–4316, Sep. 2008.
- [7] L. Dong, P. Petropulu, and H. V. Poor, "A Cross-Layer Approach to Collaborative Beamforming for Wireless Ad Hoc Networks," *IEEE Trans. Signal Proc.*, vol. 56, July 2008, pp. 2981–93.
- [8] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1713–1727, Nov. 1994.
- [9] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

- [10] D. Gesbert, S. Hanly, H. Huang, S. Shamaï, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: a new look at interference," *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1380-1408, Dec. 2010.
- [11] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE J. Select. Areas Commun.*, vol. 30, no. 3, pp. 497 – 508, April 2012.
- [12] J. G. Andrews "The Seven Ways HetNets are a Paradigm Shift", to appear in *IEEE Comm. Mag.*, Mar. 2013.
- [13] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Spectrum Leasing as an Incentive towards Uplink Macrocell and Femtocell Cooperation," *IEEE J. Select. Areas Commun.*, vol. 30, no. 3, 617 – 630, April 2012
- [14] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, pp. 1580–1589, Sep. 2000.
- [15] L. Sanguinetti, M. Morelli, and G. Imbarlina, "An EM-based frequency offset estimator for OFDM systems with unknown interference," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4470–4475, Sep. 2009.
- [16] M. Ghogho and A. Swami, "Training design for multipath channel and frequency-offset estimation in MIMO systems," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3957–3965, Oct. 2006.
- [17] O. Besson and P. Stoica, "On parameter estimation of MIMO at fading channels with frequency offsets," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 602–613, Mar. 2003
- [18] M. O. Pun, M. Morelli, and C.-C. J. Kuo, "Maximum-likelihood synchronization and channel estimation for OFDMA uplink transmissions," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 726–736, Apr. 2006.
- [19] T. Pham, A. Nallanathan, and Y. Liang, "Joint channel and frequency offset estimation in distributed MIMO flat-fading channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 648–656, Feb. 2008
- [20] J. Chen, Y.-C. Wu, S. Ma, and T.-S. Ng, "Joint CFO and channel estimation for multiuser MIMO-OFDM systems with optimal training sequences," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4008-4019, Aug. 2008.
- [21] H. Mehrpouyan and S. D. Blostein, "Bounds and algorithms for multiple frequency offset estimation in cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1300–1311, Apr. 2011.
- [22] H. G. Jeon, K. S. Kim, and E. Serpedin, "An efficient blind deterministic frequency offset estimator for OFDM systems," *IEEE Trans on Communications.*, vol. 59, no. 4, Apr. 2011.
- [23] F. F. Gao, T. Cui, A. Nallanathan, and C. Tellambura, "Maximum likelihood based estimation of frequency and phase offset in DCT OFDM systems under non-circular transmissions: Algorithms, analysis and comparisons," *IEEE Trans on Communications.*, vol. 56, no. 9, pp. 1425–1429, Sep. 2008.
- [24] P. A. Parker, P. Mitran, D. W. Bliss, and V. Tarokh, "On bounds and algorithms for frequency synchronization for collaborative communication systems," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3742-3752, Aug. 2008.
- [25] B. Zarikoff and J. Cavers, "Coordinated multi-cell systems: Carrier frequency offset estimation and correction," *IEEE Trans. Select. Areas Commun.*, vol. 28, no. 9, pp. 1490–1501, Dec. 2010
- [26] R. Mudumbai, G. Barriac, and U. Madhow, "On the feasibility of distributed beamforming in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1754–1763, May 2007.
- [27] D. R. Brown III and H. V. Poor. "Time-Slotted Round-Trip Carrier Synchronization for Distributed Beamforming," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5630-5643, November 2008.
- [28] R. Preuss and D. R. Brown III. "Two-Way Synchronization for Coordinated Multicell Retrodirective Downlink Beamforming," *IEEE Trans. Signal Process.*, vol. 59, no. 11, 5415-5427, November 2011.
- [29] N. Varanese, U. Spagnolini, Y. Bar-Ness, "Distributed frequency-looked loops for wireless networks," *IEEE Trans on Communications.*, vol.59, n.12, pp. 3440-3451, Dec. 2011

- [30] S. M. Kay, *Fundamentals of Statistical Signal Processing Estimation Theory*, Upper Saddle River, NJ: Prentice-Hall, 1993.
- [31] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1910-1917, Nov. 2003.
- [32] L. L. He, S. D. Ma, Y. C. Wu, and T. S. Ng, "Semiblind iterative data detection for OFDM systems with CFO and doubly selective channels," *IEEE Trans. on Commun.*, vol. 58, no. 12, pp. 3491-3499, Dec. 2010.
- [33] E. Simon, L. Ros, H. Hijazi, and M. Ghogho, Joint carrier frequency offset and channel estimation for OFDM systems via the EM algorithm in the presence of very high mobility," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 754-765, Feb. 2012
- [34] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498-519, Feb. 2001.
- [35] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology," *Neural Computation*, vol. 13, no. 10, pp. 2173-2200, 2001.
- [36] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in Gaussian graphical models," *Journal of Machine Learning Research*, vol. 7, Oct. 2006.
- [37] M. Leng and Y.-C. Wu, "Distributed clock synchronization for wireless sensor networks using belief propagation," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5404-5414, Nov. 2011.
- [38] A. Papoulis and S. U. Pillai, *Random Variables and Stochastic Processes*, 4th ed. New York: McGraw-Hill, 2002.
- [39] S. Puntanen, G. Styan, and J. Isotalo, *Matrix Tricks for Linear Statistical Models Our Personal Top Twenty*, Springer, 2011.
- [40] A. M. Sarhan, N. M. El-Shazly, and E. M. Shehata, On the existence of extremal positive definite solutions of the nonlinear matrix equation  $\mathbf{X}^r + \sum_{i=1}^m \mathbf{A}_i^* \mathbf{X}^{\delta_i} \mathbf{A}_i = \mathbf{I}$ ," *Mathematical and Computer Modelling*, vol. 51, no. 9- 10, 2010.
- [41] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [42] J. W. Demmel *Applied Numerical Linear Algebra*. SIAM, 1997.

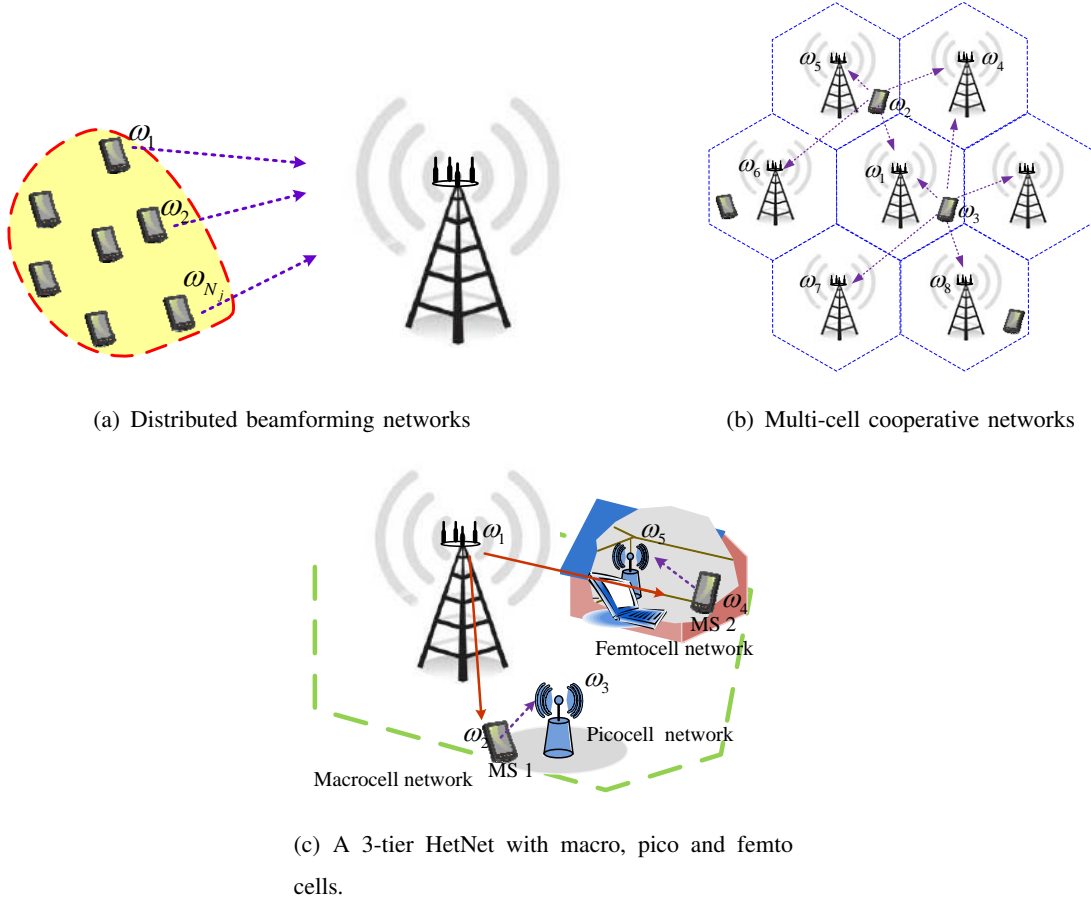


Fig. 1. Scenarios that need network-wide frequency synchronization

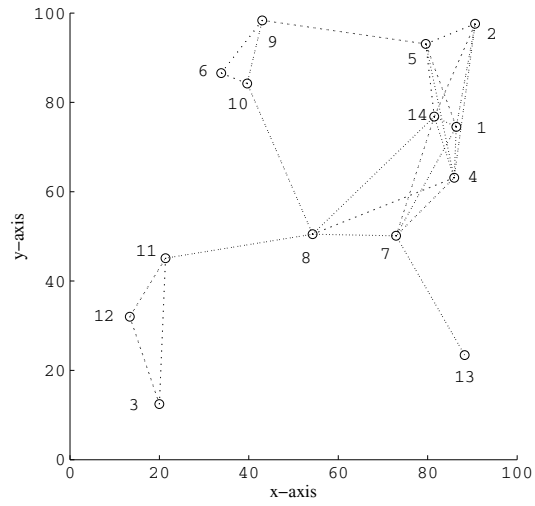


Fig. 2. An example of arbitrary network topology with 14 nodes.

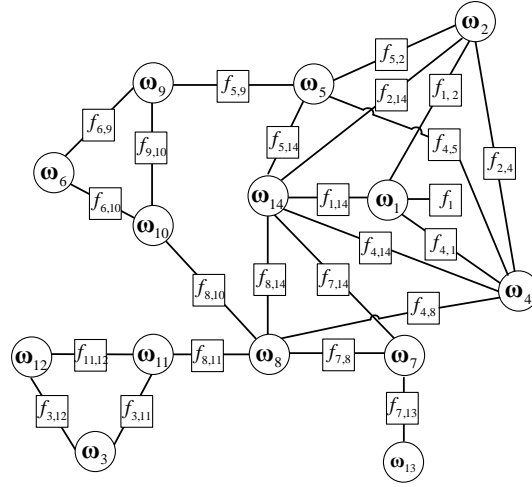


Fig. 3. The factor graph representation of the network in Fig. 2.

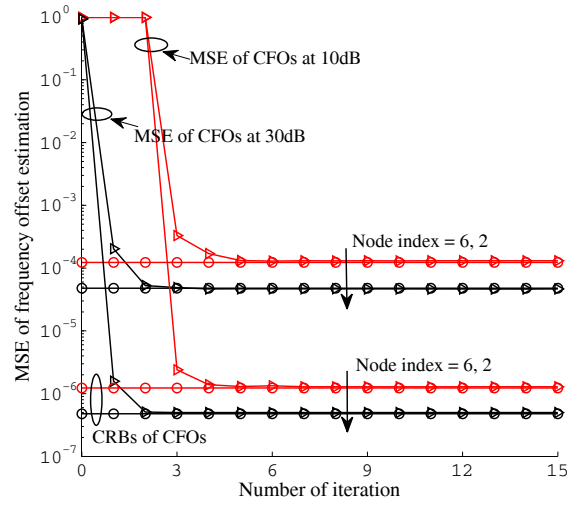


Fig. 4. Convergence performance of the proposed algorithm at different nodes.

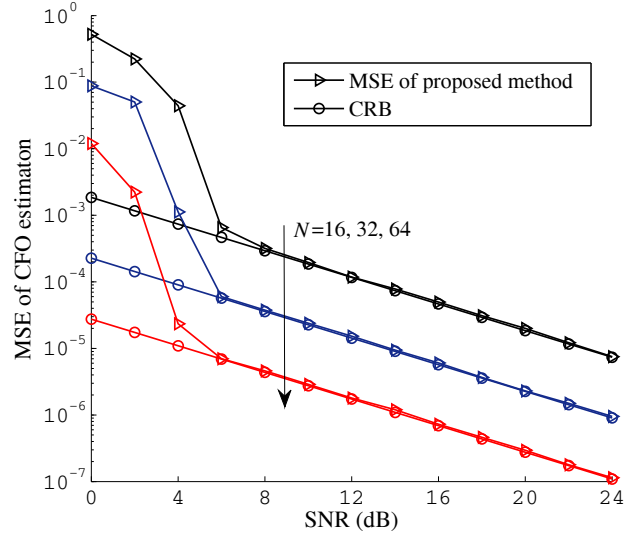


Fig. 5. MSE of CFOs  $\{\omega_i\}_{i \in \mathcal{V}}$  averaged over the whole network with respect to SNRs.

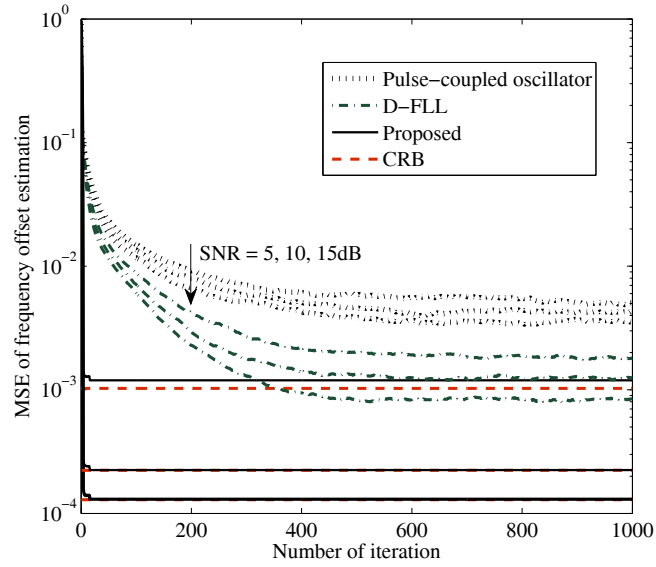


Fig. 6. Convergence of the proposed method and consensus method in [29] in single antenna case.